

人工智能概述

人类社会的每一次技术进步，都会催生新的生产、生活方式与社会思潮。历次工业革命给经济、政治、文化领域带来的系统性变革已是不争的事实。在以智能制造为主导的第四次工业革命中，人工智能无疑是对其起到巨大推动作用的重要技术。

或许是科幻作品中的人工智能动辄拥有自我意识，随时准备反攻人类，让普罗大众对其产生了遥远而不切实际的印象，实际上，人工智能早已渗透进我们日常生活的方方面面——从手机上的语音助手、美颜相机、翻译软件，到各类网站及 App 中的个性化推荐，从家居生活中的智能音箱、扫地机器人，到门禁考勤、公交安检中的人脸识别——人工智能技术已经有了不少成熟的应用，也吸引了越来越多的人才自发投入其中。而纵观全球，美国、中国、欧盟、英国、德国、法国、日本等诸多国家和国际组织都发布了人工智能的相关战略或规划，人工智能的发展势头锐不可当。

学习人工智能、研究人工智能，既是实现个人价值的需要，也是顺应时代发展的需要。

1.1 什么是人工智能

1.1.1 人工智能的定义与分类

概念的界定是一切研究展开的前提。什么是人工智能？这是每一本关于人工智能的著作都必须首先解答的问题。

追本溯源，“人工智能”一词首次正式出现是在 1955 年 8 月 31 日的《人工智能达特茅斯夏季研究项目提案》中。参与此提案的学者麦卡锡 (J. McCarthy)、马文·明斯基 (M. L. Minsky)、罗切斯特 (N. Rochester)、香农 (C. E. Shannon) 这样写道：“我们提议 1956 年暑期在新罕布什尔州汉诺威的达特茅斯学院进行一次为期 2 个月、10 人参加的人工智能研究。该研究是基于这样一种猜想进行



的，即学习的每个方面或智能的任何其他特征在原则上都可以被如此精确地描述，以至于可以制造机器来模拟它。我们将尝试寻找如何使机器使用语言，形成抽象和概念，解决现在留给人类的各种问题，并改进自己。”提案中还有一个更明确的定义：“就目前的目的而言，人工智能问题被认为是使机器以一种被称为智能的方式运行，如果人类如此表现的话。”更直白地说，他们认为人工智能就是用机器模仿人的智能。

这是对人工智能较早且较为流行的一个定义。但这种定义方式的背后其实是仿生学的思路，它没有考虑到机器产生非类人的智能的可能，也就是说，机器有可能通过与人类智能完全不同的形式达到我们所谓智能的效果。这方面最有名的例子就是 AlphaGo Zero——谷歌旗下 DeepMind 公司推出的围棋人工智能程序。它在学习围棋的过程中完全抛弃了人类棋手的经验，通过自我训练的方式，以 100 : 0 的战绩战胜了此前打败过人类顶尖棋手李世乭、柯洁等人的 AlphaGo。

麦卡锡后来也认识到了这种定义方式的局限，他在 2007 年一篇写给外行的人工智能科普文章中给出了新的释义：“它是制造智能机器，特别是智能计算机程序的科学和工程。它与使用计算机理解人类智能的类似任务有关，但人工智能不必局限于生物学上可观察的方法。”他还进一步对智能作出了解释：“智能是实现世界上目标的能力的计算部分。人类、许多动物和一些机器都具有不同种类和程度的智能。”但是，只有计算才是智能吗？很显然，人与环境的交互并不都是由计算来驱动的，虽然计算对智能机器而言必不可少。

从这个角度来说，尼尔斯·尼尔森（Nils J. Nilsson）对智能及人工智能的看法或许更为合适：“人工智能是致力于使机器变得智能的活动，而智能是使实体能够在其环境中适当地运作并具有远见的品质。”类似地，斯图尔特·罗素（Stuart Russell）和彼得·诺维格（Peter Norvig）在人工智能教材《人工智能：一种现代的方法》中给出了这样的定义：“我们将人工智能定义为从环境中接受感知并执行行动的智能体（agent）的研究。”他们还进一步地将人工智能分为四类：像人一样行动、像人一样思考、合理地思考、合理地行动。

中国电子技术标准化研究院发布的《人工智能标准化白皮书（2018 版）》中也尝试给出了自己的定义：“人工智能是利用数字计算机或者由数字计算机控制的机器模拟、延伸和扩展人的智能，感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。”

可以看到，同其他难以界定的科学概念一样，学界并没有在“人工智能”的定义上达成共识。事实上，在人工智能的研究成熟之前，我们很难给出一个



准确而适当的定义。在理论上或实践中追求具体而明确的目标无疑更具有现实意义。不过，了解这些行业先驱与巨擘的观点，可以帮助我们更好地理解这个方兴未艾的人工智能世界。

与定义上的诸多分歧不同，在人工智能的类型问题上，学者们的意见比较一致。根据能力的不同等级，人工智能可以分为两类：弱人工智能（weak AI）和强人工智能（strong AI）（图 1-1）。



图 1-1 人工智能的类型

弱人工智能，也称狭义人工智能（narrow AI）或专用人工智能（artificial narrow intelligence, ANI），简单来说就是达到专用或特定技能的智能。我们目前能够成功实现和应用的人工智能都属于弱人工智能——哪怕是看起来很厉害的无人驾驶汽车和 AlphaGo——因为它们擅长的都只是单一的任务，无法在多个领域发挥作用。

强人工智能，也称通用人工智能（artificial general intelligence, AGI），是指达到或超越人类水平的、能够自适应地应对外界挑战的、具有自我意识的人工智能。也有学者对此进行了细分，把达到人类水平的称作强人工智能，而超越人类水平的则是超人工智能（artificial super intelligence, ASI）。无论是强人工智能还是超人工智能，目前都还停留在幻想的阶段。但是，根据大多数人工智能专家的看法，超越人类能力的人工智能一定会出现。因为比起人类相对固定的智能水平，机器的智能正随着算法的优化、处理能力的增强和内存的增加而快速增长，机器超越我们只是时间问题。超人工智能到来的时刻被称作奇点（singularity）。对于奇点何时出现的问题，人工智能科学家也多次调查过同行们的看法，大多数人预计的时间都是 2060 年之前。这个估计会太过乐观吗？让我们一起拭目以待。

如果从发展的眼光来看，人工智能又可以分为运算智能、感知智能、认知智能和自主智能，这也是人工智能应用的四个不同层次（图 1-2）。

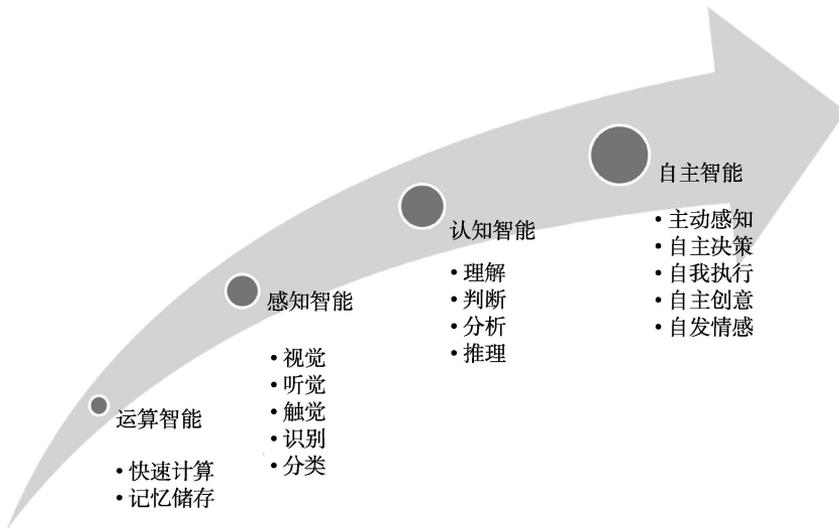


图 1-2 人工智能的四个层次

（1）运算智能，即快速计算和记忆存储的能力，这也是计算机的核心能力。

（2）感知智能，即视觉、听觉、触觉、识别、分类的能力。人类和高等动物都是通过自身丰富的感觉器官，获取环境信息，与外界进行交互的。目前在机器人身上应用的各种传感器和语音、图像、视频识别与分类等技术就是感知智能的体现。整体来说，运算智能和感知智能还停留在工具层面，并没有触及智能的核心。

（3）认知智能，即理解、判断、分析、推理的能力。而现阶段的人工智能虽然在运用自然语言处理、知识图谱、深度学习（deep learning）机制和神经网络（NNs）后，做到了一定程度的“能理解、会思考”，但仍然非常有限。此外，人类情绪对认知的影响，乃至作为认知主要部分的潜意识，都是目前机器的认知智能难以模仿实现的。

（4）自主智能，即主动感知、自主决策、自我执行、自主创意、自发情感的能力。这种自主，不仅仅是无须人类干预就可以自由移动并与人类和其他物体交互——目前的无人机、无人驾驶等技术已经实现了某种程度上的自主，更重要的是拥有自我意识、自我认知乃至自我价值观——这是目前只存在于科幻小说与电影中而现实的人工智能尚未或许永远也无法触及的部分。

1.1.2 人工智能的起源与发展

回顾历史，人工智能其实可以追溯到古代的“人造人”想象。在 2700 年



前的古希腊神话中，就出现了塔罗斯（Talos）、潘多拉（Pandora）、“黄金女仆”（the Golden Maidens）和加勒提亚（Galatea）这四个人造人的形象。前三者都是火与工匠之神赫菲斯托斯（Hephaestus）制造的：塔罗斯是用青铜铸造的巨人，他受命守卫着克里特岛，防止外来者的侵犯；潘多拉作为对普罗米修斯盗火的惩罚，其身体由黏土塑造，被众神赋予了诱人的魅力、语言的技能以及装满了灾厄的魔盒；黄金女仆则是用黄金锻造的女机器人，她们会开口说话并协助赫菲斯托斯在其宫殿中进行高难度的工作。加勒提亚则是塞浦路斯岛的国王皮格马利翁（Pygmalion）用白色象牙雕刻出的理想女性，她被爱神阿佛罗狄忒（Aphrodite）赋予了生命。不止西方，在中国战国时期的典籍《列子》汤问篇中，也记载有西周时期的巧匠偃师向周穆王进献能歌善舞、以假乱真的人偶的故事。可以看出，这些神话传说中的人造人拥有的能力与人们如今对人工智能的期许并无多大不同，尽管这些能力的实现基本上靠的是神力而非人力。

虽然古代人并不相信自己能像神一样造人，但他们在实际的生产生活中制造了许多减轻或替代人类劳动的工具，如耕地的犁、翻地的耙、灌溉的水车、收割的镰刀等。到了中世纪，人们开始制造自动机械装置，即自动机（automaton）。已知最早有据可考的自动机来自阿拉伯的博学家艾尔-加扎利（Al-Jazari），他发明了一艘载有 4 个木偶的小船，可以通过水流的驱动让木偶演奏音乐。因为互换负责音符的木栓可以让木偶演奏出不同的旋律，所以这被认为是第一个可编程的人形机器人。14 世纪，机械钟出现，钟表业开始发展。发条、齿轮等钟表技术渐渐被扩展开来用于制造机械动物和人偶，如意大利博学家达·芬奇的机械狮和机器武士。到了 18 世纪，随着第一次工业革命的逐步展开，人类从手工劳动进入大机器生产的时代。机器生产的发达，使得更多的自动机开始出现，如法国发明家雅克·德·沃康松（Jacques de Vaucanson）的消化鸭，匈牙利发明家沃尔夫冈·冯·肯佩伦（Wolfgang von Kempelen）的土耳其行棋傀儡（the Turk），瑞士钟表匠皮埃尔·雅克-德罗兹（Pierre Jaquet-Droz）的三个自动机械人偶“小作家”“小绘图师”和“小音乐家”等。虽然这些自动机只是社会上层娱乐的玩物，但作为模仿生物及人类智能行为的机器，它们可以被视作人工智能的前导和先声。

从制造工具到制造机械人偶，这一变化不仅意味着人的工具职能更多地被机器替代，更昭示了一种世界观的转变——人对自我的认识从“人是上帝的造物”变成了“人是机器”。17 世纪初，法国哲学家、数学家笛卡儿（René Descartes）提出了身心二元论，认为人是由身体和心灵两种完全不同的实体组成的。同时他还认为动物的身体只不过是复杂的机器，人体功能是以机械方式



发生作用的。与笛卡儿同时代的英国政治家、哲学家霍布斯（Thomas Hobbes）虽然并不赞成笛卡儿的二元论，但他却更进一步地认为，人是一种由上帝创造的“像钟表一样用发条和齿轮运行的‘自动机械机构’”：人的“‘心脏’无非就是‘发条’，‘神经’只是一些‘游丝’，而‘关节’不过是一些‘齿轮’”。而到了18世纪，作为机械唯物主义代表的法国思想家、哲学家拉美特利（Julien Offroy de la Mettrie）则完全抛弃了上帝，并且非常直白提出了“人是机器”的观点。他认为，人的心灵活动依赖于大脑和整个身体组织，因此和身体活动一样，也属于机械运动。所以，人整个就是一台机器。虽然这种把思想当作物质属性来论证的方式缺乏说服力，但是将人类身体机械化的观点却影响深远。在现代科技的支持下，人类的部分身体组织已经可以被仿生义肢、机械外骨骼等机械装置替代或者增强，各种模仿人体形态和行为的仿人机器人也不断涌现。

但是，身体的机械化还不足够，人工智能的思想根源在于人类心灵（或者说人类思维活动）的机械化。这一观点也可以追溯到笛卡儿。笛卡儿将数学提升为一种普遍适用的科学方法，提出了“普遍数学”，即把数学最一般的特征“度量”和“顺序”运用到其他学科来认知万物。他认为，人类认识领域的任何问题都可以转化为数学问题，人类的认识过程就是数学计算。类似地，霍布斯也认为，真正科学的知识只有在感觉经验的基础上运用推理方法才能获得，而推理就是计算。到了功利主义学派代表人物边沁（Jeremy Bentham）那里，计算就不仅限于心灵的认知层面了，人的情感、欲求、感受等的产生都是基于心灵对快乐和痛苦这两种体验的程度的计算与比较，而趋乐避苦是人类一切思想、情感、行为的动机。所以，心灵的本质就是计算。

然而，由于各种生理因素的限制，人类心灵计算的能力参差不齐，并不完善。于是，人们就开始了将数学运算机械化的尝试，企图通过机器的运算来实现纯粹的、完美的数学运算。1642年，法国哲学家、数学家帕斯卡（Blaise Pascal）发明了第一台机械计算器——加法器（Pascaline）。1673年，德国哲学家、数学家莱布尼茨（G.W. Leibniz）发明了第一台机械式的十进制四则运算器。数学运算的机械化就此开始并不断发展，直到现在的电子计算机阶段。当然，电子计算机的研制成功以及后来人工智能的诞生，还离不开数理逻辑的发展——这是一门用数学的方法来研究形式逻辑，以及研究形式逻辑在其他数学领域的应用的学科。从布尔的布尔代数、弗雷格的一阶谓词演算系统，到哥德尔的不完全性定理、克林的一般递归函数理论，再到图灵（Alan Turing）的理想计算模型图灵机，这些经典的数理逻辑理论成果为1956年人工智能的正式诞生奠定了坚实的基础。



1956年7月到8月，麦卡锡召集的人工智能夏季研讨会在达特茅斯顺利进行。会议聚集了当时相关领域的顶尖研究人员，对人工智能的问题展开了开放式的讨论。尽管从结果来看，与会人员并未就人工智能领域的标准方法达成一致，甚至对“人工智能”这个叫法都存在分歧，但他们都认同人工智能是可以实现的。这次会议也催化了之后蓬勃发展的的人工智能研究，因此被后人视为人工智能诞生的标志。

不过，在1956年之前，人工智能的相关研究已经开始了。1943年至1955年这段时期可以称作人工智能的孕育期。1943年，沃伦·麦卡洛克（Warren McCulloch）和沃尔特·皮茨（Walter Pitts）的人工神经元模型应该是现代人工智能领域最早的研究成果。1951年，马文·明斯基与同学迪恩·爱德蒙（Dean Edmunds）合作建造了世界上第一台神经网络计算机SNARC（随机神经模拟强化计算器）。当然，这一时期最重要的里程碑事件还是1950年英国数学家艾伦·图灵提出的图灵测试——如果一台机器能够与人类展开对话（通过电传设备）而不能被辨别出其机器身份，那么称这台机器具有智能。虽然严格来说，图灵测试并不严谨、完善，但其中已经蕴含有人工智能的原始概念，并明确了人工智能未来的发展目标，其划时代的意义是毋庸置疑的。

从1956年开始，人工智能的发展大致经历了三大阶段：第一阶段，从1956年到1979年，这是人工智能的诞生时期；第二阶段，从1980年到2010年，人工智能开始步入产业化；第三阶段，从2011年至今，人工智能的研究和应用迎来爆发。当然，这几十年的发展历程并非一路高歌猛进，也曾经出现过几次起伏。所以，这三个阶段又可以细分为六个时期（图1-3）。

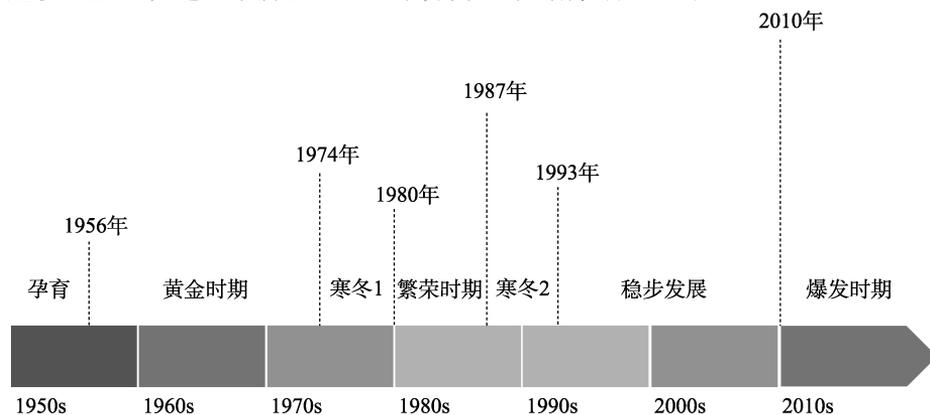


图 1-3 人工智能发展的时间轴

1. 1956—1974年，黄金时期

达特茅斯会议后的近20年是人工智能发展的黄金时期，研究者们普遍乐



观，对人工智能的热情和期望很高。这一时期代表性的研究成果有：感知器被发明，人工神经网络 [也称连接模型 (connection model)] 迎来了第一次热潮；麦卡锡开发了编程语言 LISP，这是人工智能研究中最流行且仍受青睐的编程语言；工业机器人 Unimate 被部署在美国通用汽车公司，代替人类进行危险的装配工作；世界上第一个聊天程序 ELIZA 诞生，它可以用英语与人交流；第一个可自主移动的机器人 Shakey、第一个人形智能机器人 WABOT-1 诞生；第一个专家系统^①DENDRAL 研究完成并投入使用，它的作用是帮助化学家判断特定物质的分子结构。

2. 1974—1980 年，第一次寒冬

尽管如此，黄金时期的很多乐观承诺并没有如期兑现，人们开始对人工智能产生怀疑。1973 年，应用数学家詹姆斯·莱特希尔 (James Lighthill) 为英国科学委员会编写的一份人工智能研究现状报告发表。该报告对人工智能研究的许多核心方面都作出了非常悲观的预测，称“迄今为止，该领域的任何发现都没有产生当时承诺的重大影响”。莱特希尔报告直接导致英国及其他各国政府和机构减少了对人工智能研究的资金投入，人工智能的发展进入第一次寒冬。

3. 1980—1987 年，繁荣时期

然而，寒冬中也孕育着生机。1978 年，美国卡内基梅隆大学开始为 DEC 公司研发一款能制定计算机硬件配置方案的专家系统 XCON。1980 年，XCON 投入商业使用，为 DEC 公司节省了大量成本。XCON 的商业成功吸引了许多公司的效仿，专家系统所依赖的知识处理问题也成为这一时期的研究焦点。1982 年，日本推出了第五代计算机计划，其目标是造出具有人工智能的计算机系统。随后，美、英、法、德、苏联等国也纷纷响应，投入资金加入角逐，人工智能开始进入新一轮的发展。

4. 1987—1993 年，第二次寒冬

20 世纪 80 年代末期开始，个人电脑的性能不断提升冲击着 AI 硬件市场，曾经大获成功的专家系统暴露出应用领域狭窄、知识获取困难、维护费用居高不下等问题，日本人宏伟的第五代计算机计划也宣告失败。各国政府和投资者再次停止为人工智能研究提供资金，人工智能进入第二次寒冬。

^① 专家系统 (expert system) 是以知识库和推理机为核心，在特定领域具有专家水平解决问题能力的智能程序系统。



5. 1993—2010 年，稳步发展

进入 20 世纪 90 年代，随着计算机硬件的发展，人工智能终于取得了突破性的成果。这一时期最重要的里程碑事件莫过于 1997 年 IBM 的深蓝击败了国际象棋世界冠军加里·卡斯帕罗夫，人工智能从此进入大众视野。2006 年，杰弗里·辛顿（Geoffrey Hinton）提出了深度学习的概念，英伟达（Nvidia）推出了并行计算平台和编程模型 CUDA（统一计算设备架构）。2007 年，李飞飞启动 ImageNet 项目，试图构建一个更好的数据集（data set）。研究者在算法、算力和数据三方面的努力为人工智能接下来的爆发式发展打下了基础。

6. 2011 年至今，爆发时期

2011 年以来，深度学习算法开始在人工智能的子领域广泛应用。这一时期的重要事件有：2011 年，IBM 的 Watson 在智力问答节目中获胜。同年，苹果公司的智能语音助手 Siri 问世。2014 年，亚马逊正式发布了智能音箱产品 Echo。Siri 和 Echo 引得各家厂商纷纷效仿，纷纷推出了自己的同类产品抢占市场。2015—2017 年，谷歌 DeepMind 的 AlphaGo 不断击败数位人类顶尖围棋棋手。2018 年，谷歌发布的 BERT 模型在自然语言处理领域取得了重大突破……^①如今，人工智能的核心技术不断发展，应用场景逐渐丰富，市场规模持续扩大，已成为时代发展的重要驱动力。

在人工智能的发展过程中，不同学科背景的学者对人工智能的看法各异，因而也产生了人工智能的不同学派。

1. 符号主义学派

首先要提的便是符号主义学派。符号主义（symbolicism）又称逻辑主义（logicism）、心理学派（psychologism）或计算机学派（computerism）。代表人物有艾伦·纽厄尔（Allen Newell）、赫伯特·西蒙（Herbert A. Simon）、尼尔斯·尼尔森等。

符号主义的思想源头是 19 世纪中叶出现的数理逻辑。符号主义认为人类认知和思维的基本单元是符号，而认知过程就是符号上的操作运算。人被视为一个物理符号系统，计算机也是如此。因此，计算机可以用来模拟人类的行为。他们还认为，知识是信息的一种形式，是智力的基础。人工智能的关键问题是知识表示和知识推理。概括来说，就是利用物理符号系统假设和有限合理性原理来实现人工智能。

^① 人工智能发展 70 年来的大事年表详见附录。



符号 AI (symbolic AI, sAI), 或者有效的老式人工智能 (good old-fashioned artificial intelligence, GOF AI), 在人工智能发展的早期一直占据着主流, 为人工智能的发展作出了许多重要贡献。老式并不是对其过时的指责, 而是意味着经典。只要是规则清晰、目标明确的任务, 使用符号主义的方式是非常方便的。事实上, 目前绝大多数的计算机程序和系统也还都是基于编程规则创建的。20 世纪 80 年代流行的专家系统就是符号 AI 的代表性成果之一。

尽管如此, 符号 AI 也有着非常明显的局限。它严重依赖已经设定好的规则, 并且无法处理存在大量变化的、非结构化的数据。举例来说, 让符号 AI 在图像中识别人脸就非常困难, 因为你无法穷举人脸的面貌与所处的环境, 更难以创建对应的识别规则。专家系统从 20 世纪 90 年代开始逐渐遇冷也是如此, 它所依赖的复杂符号与大量规则需要耗费大量人力, 不便于维护, 并且可以应用的领域也非常狭窄, 没有普适性。

2. 连接主义学派

虽然符号主义学派在人工智能发展的早期占据了主流, 但现下更受欢迎的却是另一个学派——连接主义 (connectionism)。连接主义又称仿生学派 (bionicsism) 或生理学派 (physiologism)。代表人物有沃伦·麦卡洛克、沃尔特·皮茨、约翰·霍普菲尔德 (John Hopfield)、鲁梅尔哈特 (D.E. Rumelhart) 等。

连接主义的思想源头是仿生学中对人脑模型的研究, 尤其是对人脑学习和记忆的研究。连接主义认为智能活动是由大量简单单元通过复杂的相互连接后并行运行的结果。人脑不同于计算机, 应该用人脑模式代替计算机模式。神经网络及神经网络间的连接机制与学习算法是这一学派的理论基础。

连接主义的发轫其实很早, 1943 年, 沃伦·麦卡洛克和沃尔特·皮茨就发表了一篇关于神经网络和自动机的论文, 对连接主义 AI 的研究影响深远。20 世纪 60 年代, 连接主义的研究也曾出现过热潮。但是由于效率低下的缺陷和对大量计算资源的需求, 人们对连接主义的兴趣逐渐降低。直到 20 世纪 80 年代, 随着基于递归的新一代神经网络、多层感知机和神经网络反向传播算法的提出, 连接主义才出现复兴。进入 21 世纪, 其更是掀起了深度学习的热潮, 在计算机视觉、自动语音识别 (automatic speech recognition, ASR)、自然语言处理等方面都取得了很大的进展, 成为当下人工智能的主流。

当然, 连接主义 AI (connectionist AI, cAI) 也并非没有缺点, 需要大量高质量的数据, 算法缺乏透明度, 难以进行合理论证都是连接主义 AI 一直被诟病的问题。此外, 连接主义 AI 也很难解决需要逻辑和推理的任务, 而这



恰恰是符号 AI 擅长的。

3. 行为主义学派

除了符号主义和连接主义，还有学者将行为主义（actionism）算作第三种学派。行为主义又称进化主义（evolutionism）或控制论学派（cyberneticsism）。代表人物有诺伯特·维纳（Norbert Wiener）、罗德尼·布鲁克斯（Rodney Brooks）等。

行为主义学派的兴起源于控制论。1948 年，诺伯特·维纳出版了《控制论——或关于在动物和机器中控制和通信的科学》，标志着控制论这门学科的诞生。控制论研究的是生命体、机器和组织的内部或彼此之间的控制与通信。控制论中的智能性原则认为不仅在人类和人类社会中，在其他生物群体乃至无生命的机械世界中，都存在着同样的信息、通信、控制和反馈机制，智能行为是这套机制的外在表现，因此不仅人类，其他生物甚至是机器也同样能做出智能行为。行为主义 AI 以此为理论基础，提出了“感知—行动”的智能行为模拟方法，认为人工智能可以像人类智能一样，在与周围环境的交互过程中通过反馈机制不断进化，发展出越来越强的智能。

由于控制论的原因，不少学者并不把行为主义 AI 划入人工智能的范畴。在他们眼中，沿着这一理论路径最多只能实现完美的机械自动化，难以达到真正的智能。

1.2 人工智能的技术路线

1.2.1 机器学习

在当前的人工智能领域，机器学习，或者更准确地说，“大数据 + 深度学习”的技术路线已经成为研究的主流。这种主流甚至表现在，只要对人工智能进行稍微深入一点的了解，就一定会看到“机器学习”这个名词。那么，什么是机器学习？

顾名思义，机器学习试图让机器拥有人类的学习能力。机器学习作为一个术语，最早是由人工智能领域的先驱亚瑟·塞缪尔（Arthur Samuel）在 1959 年提出的，并表示“它使计算机能够在没有明确编程的情况下进行学习”。而这种学习的目的在于改善系统的性能和效果。

这个解释可能有些抽象，让我们来举个通俗的例子。想象你站在一个篮球



场上，现在你需要完成罚球线上的定点投篮。第一次投篮，你用的力气太小，篮球甚至没有接触到篮筐。于是在第二次，你使用了更大的力气，结果这次篮球砸在篮板上弹了出去。到了第三次，你终于找到了最合适的力道，投篮命中。在这个过程中，每一次失败的投篮都是一次经验的累积，通过对这些经验的利用，你才能在最后一次投篮成功。

机器学习的过程也是类似的。计算机使用学习算法（learning algorithm）从数据中累积经验，生成模型（model）——这个过程称为“学习”（learning）或者“训练”（training），训练中使用的数据称为“训练数据”（training data）。之后面对新的情况时，模型就可以帮助计算机做出判断和行动。

具体来说，在确定好任务目标之后，机器学习的过程主要包括以下几步：收集数据——数据预处理——特征提取与处理——选择合适的学习算法进行训练，生成模型——评估模型效果——调整训练过程中的参数、变量，优化模型效果——开始使用。当然，不同的机器学习方式在具体执行过程中采用的方法是有所差别的。根据学习的过程中是否有人类监督，机器学习可以分为监督学习（supervised learning）、无监督学习（un-supervised learning）、半监督学习（semi-supervised learning）和强化学习（reinforcement learning）。

监督学习使用人类事先打好标签的数据集训练模型，然后根据新输入的数据，来判断它的类别，或者预测它的值，也就是所谓的分类和回归。例如，判断你今晚看到的月亮是阴是晴是圆是缺就是分类，而预测未来的月相就是回归。监督学习是机器学习中最常用的类型，其经典算法模型有 K-近邻（KNN）、线性回归、朴素贝叶斯、逻辑回归、决策树（decision tree）、支持向量机（support vector machine, SVM）、支持向量回归（SVR）、隐马尔可夫模型（HMM）、条件随机场（CRF）等。

而在无监督学习中，机器学习算法需要在未标注的数据中寻找模式，因为有的时候人工标注的成本很高，或者人类也缺少相关经验无法进行标注。无监督学习不像监督学习拥有明确的任务目标，学习效果也无法量化评估，但可以帮助我们发现一些没有注意到的规律或者趋势。无监督学习最常见的算法是聚类和降维。聚类就是对数据进行特征提取并分类，常用的算法有 K 均值聚类（K-means）、层次聚类、谱聚类、EM（最大期望）算法、高斯混合模型（GMM）等。降维有点类似压缩，它是使用更少的但更有效的特征来表示数据，常用的算法有主成分分析（PCA）、奇异值分解（SVD）、因子分析（FA）、*t* 分布随机近邻嵌入（*t*-SNE）等。无监督学习在异常检测、网购推荐中都有所应用。

半监督学习则是综合了监督学习和无监督学习的一种方法，它使用少量的



有标签数据和大量的无标签数据进行训练，可以解决监督学习对数据标注的依赖问题，也可以解决无监督学习准确率较低、应用范围有限的问题。半监督学习对无标签数据的使用依赖于半监督假设，当假设正确时，半监督学习才能实现较好的学习性能。半监督学习的三大基本假设包括：①平滑假设 (smoothness assumption)：彼此更接近的点更有可能具有相同的类别标签。②聚类假设 (cluster assumption)：数据可以分为离散的集群，同一集群中的点更有可能具有相同的类别标签。③流形假设 (manifold assumption)：将高维数据嵌入低维流形中，当两点位于低维流形中的一个小局部领域内时，它们具有相同的类别标签。因为是对监督学习和无监督学习的综合，所以分类、回归、聚类、降维这四种方法半监督学习都有，具体如图 1-4 所示。

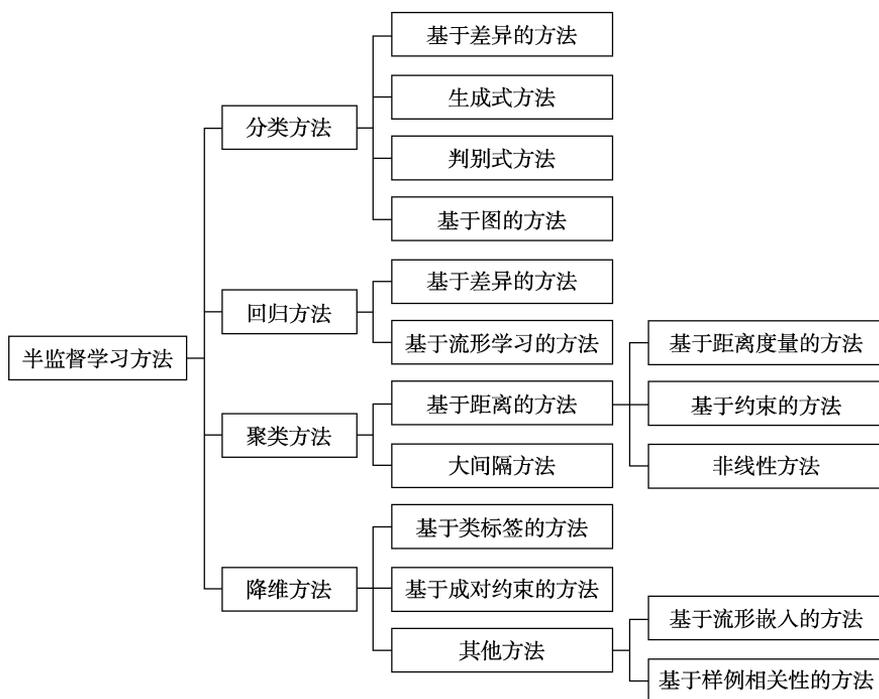


图 1-4 半监督学习方法

至于强化学习，它与上述机器学习方式的不同之处在于，它采用的是一种奖励的思路：强化学习算法并不需要依靠事先准备好的数据（无论是否有标签）来学习，而是从自身行为获得的奖励多少来累积经验，以最大回报为目标驱动结果的改进。强化学习有五个关键的要素：代理 (agent)、环境 (environments)、状态 (states)、动作 (actions) 和奖励 (rewards)。处在某个环境中的代理在执



行了某个动作后，使环境转换到了一个新的状态。环境根据这个新的状态给出奖励信号（正奖励或负奖励），代理根据接收到的奖励反馈和新的环境状态再去执行新的动作。整个过程会不断循环，直到代理获得最大的奖励。这种描述听起来很像游戏，事实上强化学习在游戏中的应用也是最多的。强化学习常用的算法有蒙特卡洛方法（Monte-Carlo learning）、时序差分学习（temporal-difference learning）、SARSA 算法、Q-learning 算法等。

近年来很火的深度学习其实也是机器学习的一种，但它并不适用于上述的分类标准，因为它是利用神经网络来模仿人类大脑的方法。深度学习的起源可以追溯到 20 世纪 40 年代，但直到 2006 年相关研究才取得较大突破，进入快速发展的时期。深度学习的核心过程其实非常简单，一共就三步：选择神经网络架构——确定学习目标——开始学习。目前比较常用的神经网络架构包括深度神经网络（deep neural networks, DNN）、时间延迟神经网络（time delay neural network, TDNN）、深度置信网络（deep belief networks, DBN）、卷积神经网络（convolutional neural networks, CNN）、卷积深度置信网络（convolutional deep belief networks, CDBN）、生成对抗网络（generative adversarial networks, GAN）、循环神经网络（recurrent neural network, RNN）等。这些神经网络架构有非常多的隐藏层（可以多达 150 个），而传统的神经网络一般只包含 2~3 个，这也是“深度”这个叫法的由来。和传统机器学习相比，深度学习也非常依赖数据，而且需要的数据量更大，两者在数据准备和预处理方面都是很相似的。它们的差别主要体现在数据特征的提取上，传统机器学习主要依赖人工，而深度学习则是靠算法自动提取的。这也是深度学习会被人诟病可解释性较差的原因。深度学习目前已经应用到了图像分类、语音识别、自然语言处理、自动驾驶汽车等各个细分领域的研究当中，当然，在大众认知中最有名的还是打败了围棋世界冠军的人工智能 AlphaGo。

1.2.2 类脑智能计算

虽然机器学习及深度学习已经成为当下人工智能领域主流的技术路线，但是其底层仍未摆脱传统的“冯·诺依曼”计算机体系架构，计算能力依然受限。如果这个载体彻底更换，可能为强人工智能的实现带来新的机会。因此部分研究者提出了类脑智能的想法，希望以生物脑作为参照，利用脑科学的研究成果，构造逼近生物神经网络的电子神经系统，推动人工智能的新发展。

目前深度学习所使用的人工神经网络从其出发点来看，可以算作类脑的尝试，但它实质上只是参考了大脑神经细胞间部分拓扑结构而搭建的数学模型，



并非来自生物神经系统的数理解析结果，因此不能算作真正的类脑。从脑科学的发展来看，人类对大脑的认知还十分有限，类脑智能的研究自然也只是处于萌芽阶段，目前的研究方向集中在类脑模型与类脑信息处理、类脑芯片与计算平台等方面。不过，各大国都意识到了脑科学研究的重要性，欧盟、美国、日本、中国等都推出了国家级的脑计划，成立了一批相关的研究机构。未来的类脑智能能否有所突破，还需要脑科学、神经科学、认知科学、计算机科学、人工智能等各领域研究者的共同努力。

1.2.3 量子智能计算

除了类脑智能，还有研究者提出将量子力学与人工智能结合起来，把量子计算作为人工智能发展的另外一条可能的技术路线。

20 世纪 80 年代，诺贝尔物理学奖获得者理查德·费曼（Richard Feynman）有感于经典计算机的局限，提出了量子计算的概念。1994 年，贝尔实验室的物理学家皮特·秀尔（Peter Shor）对量子比特的研究，证明了量子计算机在计算速度上的优势，研究者们开始构建具有更多量子比特的量子计算机。2011 年，加拿大计算机公司 D-Wave 制造出了拥有 128 个量子比特的首台商用量子计算机，量子计算机的研究开始飞速发展。这也从硬件层面促进了量子人工智能的发展。而从国家政策层面来看，欧盟出台的《量子宣言》，以及美国的《量子计算发展白皮书》也都说明了政府层面对这项技术的看好。

目前的量子人工智能主要依靠脑量子场论、神经系统的量子态论、微管引力理论作为理论基础，其研究主要还是集中在对已有机器学习、深度学习的优化上，并没有出现算法上的突破。同类脑智能一样，量子智能的发展还需要更多研究人员的投入。

1.3 人工智能的基础应用技术

在大方向的技术路线之下，人工智能存在着众多具体的应用技术。在目前弱人工智能的发展阶段，自然语言处理、智能语音技术、计算机视觉（computer vision）和知识图谱（knowledge graph）是其中比较热门的四类。

1.3.1 自然语言处理

语言是人类互相交流的方式之一。我们借助语言说话、阅读、写作，表达



自己的思想和情感。而所谓自然语言，在语言学中是与人造语言对应的一个概念，它指的是一种自然地随文化演化的语言，我们平时所说的汉语、英语、法语、日语都是自然语言的例子。人造语言则是由人特意为了某些特定目的而创造的语言，如世界语。而在计算机科学领域，无论人造与否，自然语言就是人类交流的语言。与之相对应的则是计算机语言（编程语言），一种人类使用各种形式化的逻辑符号创造出的与计算机交流的语言。而自然语言处理作为人工智能和语言学的分支，研究的是如何让计算机理解和处理人类的语言。其核心任务包括两大部分：自然语言理解（natural language understanding, NLU）和自然语言生成（natural language generation, NLG）。前者是让计算机理解我们说的话，后者是让计算机学会使用我们的话。

这是一个很难的课题。众所周知，语言的规律本来就错综复杂，人们又可以随时根据表达的需要创造新的表达方式，穷举绝无可能。而语言作为一套符号系统，它所表示的具体事物或抽象概念是人类长久以来认知世界的成果体现，这种与知识天生的关联进一步增加了研究的难度。再加上，语言的多义性和歧义性要求对语言的理解必须依靠所处的环境与上下文。这些都是自然语言处理，尤其是自然语言理解不得不面对的难点。

为了达到目标，研究者们尝试过基于规则、基于统计的方式来判断自然语言的意图，现在又引入深度学习的方法，但仍未使计算机真正地“掌握”人类的语言。尽管如此，自然语言处理还是有不少较为成功的应用。

（1）垃圾邮件过滤器：大量的垃圾邮件无疑是最让电子邮箱用户困扰的问题之一。各大电子邮件服务商曾经尝试过许多办法进行邮件过滤（如设置关键词），但效果都不理想。而引入 NLP 的方法后，过滤器通过查看并理解邮件文本内容的方式来判断垃圾邮件，准确率提升了很多。谷歌的 Gmail 就是这样做的。

（2）机器翻译：让机器替代人工成为译员就是机器翻译。现在主流的搜索引擎 Google、Bing、Yahoo、百度等都有机器翻译的功能，基本上都可以直接从网页上读取文本，将其从一种语言翻译成另一种语言，方便各语种使用者和世界交流。

（3）文本摘要：互联网带来的信息爆炸让如何迅速获取有效信息成为一个亟待解决的问题。所谓文本摘要就是使用 NLP 对各类文本进行“降维”的处理，压缩提取出精练简洁的信息，以便人们可以更快地理解它。目前文本摘要在搜索结果预览、新闻标题生成中有所应用。

（4）情绪分析：随着电商的繁荣，网购的成交量不断增长，相应地，产品



的评论数也在大幅增加。企业需要从中找到改进产品的方向，消费者则需要找到选择产品的参考。如果仅仅依靠人工筛选判断，是很难得出准确结论的。而 NLP 则可以处理大量文字信息，发现文字背后的情感倾向，满足企业和消费者的需要。

(5) 聊天机器人：聊天机器人即通过文字或语音与人类交流的程序。最早的聊天机器人 ELIZA 采用的还是模式匹配和字符替换的规则，而如今的聊天机器人则运用了 NLP 的技术，并且涵盖了文本向量化、信息提取等诸多 NLP 任务，实现了更智能的聊天效果。典型代表有苹果的 Siri、微软的小冰等。

1.3.2 智能语音技术

智能语音技术专注于感知层面的智能化，致力于让机器拥有听说能力，是最早实现落地、市场应用也最为广泛的人工智能技术。智能语音技术具体可以分为自动语音识别、语音合成 (text to speech, TTS)、声纹识别 (voiceprint recognition, VPR)、语音分类等，下面分别进行简单的介绍。

1. 自动语音识别

自动语音识别，又称语音识别，简单来说就是一种把说话内容转换成文字的技术。其流程主要包括四步：语音输入——提取特征进行编码——根据声学模型和语言模型解码——转换为文字输出。语音识别的技术框架中有三个重要的组成部分：模型训练、前端语音处理、后端识别处理 (图 1-5)。

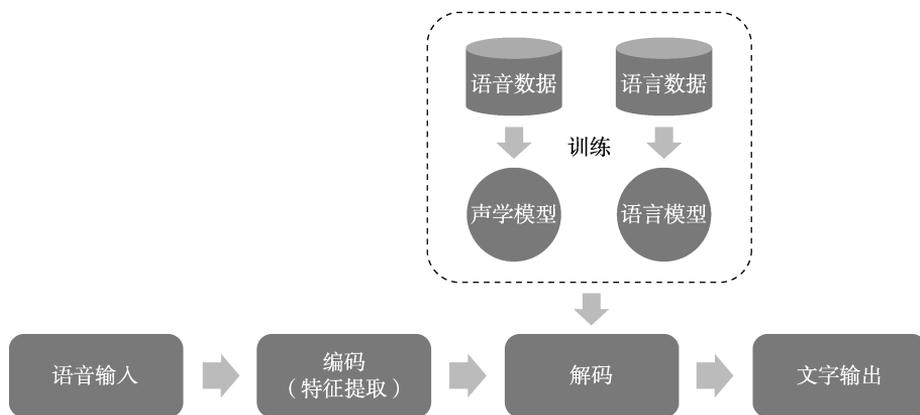


图 1-5 语音识别的原理

语音识别系统的模型通常由声学模型和语言模型两部分组成，分别对应于



从语音信号中抽取的特征到音节概率的计算和音节到字概率的计算。前端语音处理指利用信号处理的方法对说话人语音进行检测、降噪等预处理，以便得到最适合识别引擎处理的语音，并从中抽取特征向量的过程。后端识别处理就是指利用训练好的声学模型和语言模型对提取到的特征向量进行识别（也称为“解码”），得到文字信息的过程。其中最主要的解码器部分就是指对原始的语音特征进行声学模型打分和语言模型打分，并在此基础上得到最优的词模式序列的路径，此路径上对应的文本就是最终识别结果。为了得到更好的体验，还需要对识别结果进行诸如打标点、文本顺滑、数字归一化、自动文本分段等文本后处理，并将最终处理结果输出。

对于语音识别的效果，有两个评测的指标：WER（词错率）和SER（句错率）。词错率计算的是识别错误的字数占有识别字数的比例，无论是识别多了还是识别少了，都属于识别错误 [中文语音识别使用的是字错率（CER）]。句错率计算的则是识别错误的句子个数占句子总数的比例。至于怎么样才算句子识别错误，不同使用场景的标准不尽相同。有的时候只要关键词正确了就不算错，但有的时候则严格要求整句话必须一模一样。一般来说，ASR模型测试时多使用字错率作为指标，但用户体验方面的测试多使用句错率。

目前国内外业界顶尖的公司，都声称自己的语音识别准确率可以达到97%~98%的水准。但实际上，语音识别的效果受录音设备、环境噪声、混响、说话口音、说话方式、谈论话题等客观因素的影响很大。想要提高实际应用中的识别率，还是离不开对大量语音文本语料的训练。

2. 语音合成

语音合成是将数字文本转换为拟人化语音的技术。它对于外语学习者、阅读障碍者、视障人员来说都是很好的工具。

语音合成的基本流程也是四步：输入文本——进行文本分析（分词）——经过声学系统（声学模型、声码器）处理——合成语音。如图1-6所示。



图 1-6 语音合成原理

从技术层面来说，语音合成早期一直存在两条长期并存的技术路线：选音



拼接和参数合成，二者各有优劣。选音拼接即从事先录制的语音中选取所需的基本单位（音节、音素等）进行拼接。这种合成方式音质、情感真实，但字间协同、过渡生硬，对录音量和覆盖率要求高，且不能改变声线、情感，只能小幅调节语速和音高。参数合成即通过数学方法对已有录音进行频谱特性参数建模，在构建好文本序列与语音特征的映射关系后，就可以借此找到新输入文本的音频特征，然后通过声学系统生成我们能听懂的语音。这种合成方式字间协同、过渡平滑，录音量小，可变量高，声线、情感、语速和音高都可以更改，但是音质机器感浓、音色特性损失较大、缺乏临场感。

2013年以来，深度学习浪潮席卷了涉及机器学习的所有应用领域，语音合成的技术也随之更新换代为深度神经网络端到端的合成方式。所谓端到端，是把传统的文本分析和声学系统处理这两个模块合并成一个黑盒，直接输入文本就能输出合成的音频。这种合成方式减少了对语言学知识的要求，在音质、语速、流畅度、拟人度上面都有很大的效果提升。

对于语音合成的效果，目前业内一般采用 MOS 值（mean opinion score，平均意见评分）评测的方式。这是一种主观的评分方式，主要根据评分者个人对声音的喜好程度和整体感觉，进行 1~5 分的评分（可精确到小数点后一位）。其中，1 分代表极差，2 分代表差，3 分代表还可以，4 分代表好，5 分代表非常好。最后，将所有评分者的打分取平均值得到的分数就是 MOS 值。

3. 声纹识别

声纹识别，也叫作说话人识别，是通过声音特征来判断说话人身份的技术，属于生物识别技术的一种。声纹可以简单理解为声音特征，但更准确地说，是由特殊的电声转换器绘制的声波特性的频谱图案，是各种声学特征图的集合。声纹因为具有长期稳定的特征，所以被视作重要的人体身份特征用于身份认证，在金融、证券、社保、公安、军事等领域有着广泛的应用。

声纹识别的基本原理是事先为每个说话人建立一个能描述其特征的声纹模型，之后再有新的未知语音输入的时候，才能判断说话人的身份。也就是说，声纹识别的流程有两个部分：声纹注册和声纹辨认。两者都需要先输入声音，然后进行端点检测，再做特征提取，只不过前者的后续流程是声纹注册和声纹模型，而后者的后续流程是进行声纹辨认和结果判断。

声纹识别主要有两种类型：一对一的说话人确认（speaker verification）和一对多的说话人辨认（speaker identification）。说话人确认，即判断待测语音是不是某人的声音。例如在银行、证券的非柜台业务办理中，就可以用说话人确认来判断操作者是否是用户本人。说话人辨认，即判断待测语音是谁的声音。例



如一些 App 上的声纹锁功能，就是用的说话人辨认来判断当下登录的用户是多个已注册声纹用户中的哪位。声纹识别原理如图 1-7 所示。

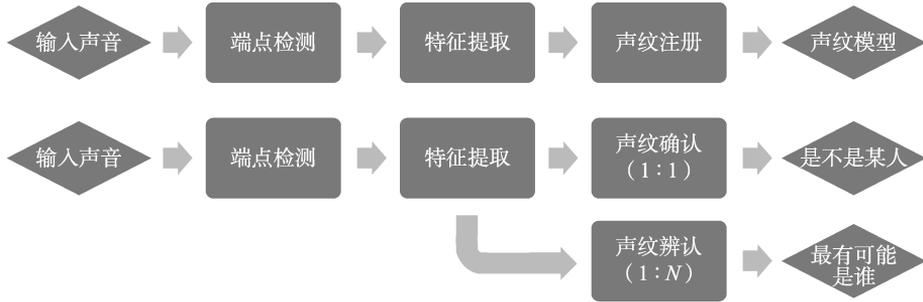


图 1-7 声纹识别原理

4. 语音分类

语音分类主要是对语音从语种、性别、年龄段等不同维度进行识别和分类。比如判断一段音频是普通话、粤语、英语还是日语，它的说话者是男性还是女性，年纪多大等。语音分类的处理流程和其他智能语音技术也很相似，包括语音输入、特征提取、特征分析、输出结果这四步，其核心算法仍然是机器学习中的各种聚类算法。

1.3.3 计算机视觉

计算机视觉是赋予机器“看”的能力的科学。它专注于创建可以像人类一样处理、分析和理解视觉数据（图像或视频）的数字系统。

那么计算机视觉是怎样让机器“看到”世界的呢？其实它的原理参考了人类视觉系统的工作方式。人类的视觉原理是：光线进入眼睛，眼睛把光线携带的信息转化成电信号传给大脑，大脑将其中关键的视觉信息提取出来进行抽象，分析其形状、运动等信息，处理完成之后就变成了视觉体验，人们也就意识到自己看到了东西。而机器的方法也是相似的：在输入视觉数据后，通过多层的神经网络，分层级一步一步地识别出图像的特征，最终通过多个层级的组合，在最后一层完成处理，输出结果。

很显然，这是一种深度学习的方法。其核心原则可以追溯到 1959 年的一项猫的视觉皮层实验——神经生理学家在实验中发现，初级视皮层的视觉处理总是从简单的结构开始，如定向的边缘。此后，计算机视觉从二维图像的分析识



别起步，逐渐出现了以理解三维场景为目的的研究、基于特征的物体识别的研究、人脸识别的研究等不同主题的研究。虽然其中也不乏光学字符识别这样的落地成果，但计算机视觉迎来爆发还是在 2012 年首个深层卷积神经网络模型 AlexNet 赢得 ImageNet 图像分类竞赛之后。这不仅得益于强大算法与硬件发展形成的合力，我们每天在互联网上生成的大量公开可用的视觉数据也是这项技术发展迅速的原因。

目前的计算机视觉主要有下列研究任务。

(1) 图像分类 (image classification): 对静止的图像内容进行分类描述, 如判断图片上的是猫还是狗。

(2) 目标检测 (object detection): 给定一张图像或是一个视频帧, 让计算机找出其中所有目标的位置, 并给出每个目标的具体类别。

(3) 语义分割 (semantic segmentation): 从字面意思上理解就是让计算机根据图像的语义来进行分割。它将整个图像分成像素组, 然后对像素组进行标记和分类, 试图在语义上理解图像中每个像素是什么。

(4) 实例分割 (instance segmentation): 目标检测和语义分割的结合, 它需要在图像中将目标检测出来, 然后给每个像素打上标签。

(5) 视频分类 (video classification): 对一个由多帧图像构成的、包含语音数据、包含运动信息等的视频对象进行分类描述。

(6) 人体关键点检测 (human keypoints detection): 又称人体姿态估计 (pose estimation), 主要检测人体的一些关键点, 如关节、五官等, 通过人体关键节点的组合和追踪来识别人的运动与行为, 描述人体姿态, 预测人体行为等。

(7) 场景文字识别 (scene text recognition): 将自然场景图片中的文字信息识别出来。和传统的光学字符识别不同, 自然场景中文字展现的形式非常丰富, 识别难度也更大。

(8) 目标跟踪 (object tracking): 在特定场景跟踪某一个或多个特定感兴趣对象的过程。

目前, 计算机视觉已广泛应用于众多行业, 从交通、安防到医疗、零售、制造业等, 且其市场还在不断扩大。预计到 2023 年, 计算机视觉的市值将达到 96.2 亿美元。

1.3.4 知识图谱

知识图谱最早是谷歌在 2012 年提出的一个概念。谷歌的知识图谱是谷歌使用从各种来源收集的信息来增强其搜索引擎结果的知识库。在使用谷歌进行



搜索时，搜索结果页面的右边会出现“知识面板”，展示关于搜索主题的汇总信息。这样用户就可以直接查看到这些信息内容，而不用一个个点开其他网站自己做汇总。

虽然知识图谱这个概念的提出并不算早，但其发展其实可以追溯到 20 世纪 60 年代符号主义学派提出的一种知识表示方法——语义网络（semantic networks）。20 世纪 80 年代，哲学概念“本体”（ontology）又被引入人工智能领域来刻画知识。1989 年，蒂姆·伯纳斯·李（Tim Berners-Lee）在欧洲高能物理研究中心发明了万维网，人们可以通过链接把自己的文档链入其中。在万维网概念的基础上，蒂姆在 1998 年又提出了语义网（Semantic Web）的概念。与万维网不同的是，链入语义网的不只是网页，还包括客观实际的实体（如人、机构、地点等）。2006 年，蒂姆又强调语义网的本质是要建立开放数据之间的链接。再然后就是谷歌发布的基于知识图谱的搜索引擎了。后来，这个概念渐渐普及，成为人工智能一项重要的应用技术。

那么，人工智能领域的知识图谱究竟是什么？知识图谱可以解释为用图谱来表示的知识库。其中，知识库存储了可供计算机用来解决问题的信息或数据，这些信息或数据代表了关于世界的事实，即知识。而图谱则是一种由一些节点和边互相连接而成的结构。更明确地说，知识图谱旨在描述真实世界中存在的各种实体或概念及其关系，一般用“实体—关系—实体”或“实体—属性—属性值”的三元组表示，比如：“姚明—夫妻—叶莉”“姚明—身高—2.26 米”。多个实体间通过关系互相连接，形成了网状的知识结构。

知识图谱按照功能和应用场景可以分为通用知识图谱与行业知识图谱。通用知识图谱面向的是通用领域，强调知识的广度，形态通常为结构化的百科知识，针对的使用者主要为普通用户；行业知识图谱则面向某一特定领域，强调知识的深度，通常需要基于该行业的数据库进行构建，针对的使用者为行业内的从业人员以及潜在的业内人士等。通用知识图谱中的知识，可以作为行业知识图谱构建的基础；而构建的行业知识图谱，也可以再融合到通用知识图谱中。通用知识图谱的广度和行业知识图谱的深度相互补充，将形成更加完善的知识图谱。

知识图谱最初因搜索而生，搜索自然也成为知识图谱目前最主要的应用场景之一。除此之外，知识图谱还被用于人机交互问答，提高机器人的知识水平，以及辅助进行数据分析与决策等。



1.4 人工智能的市场应用

在错综复杂的国际环境下,人工智能逐渐形成了由软硬件支撑层、产品层、应用层三层堆积起来的架构,它已经不仅仅停留在看不见、摸不着的代码层面,更繁衍出了各式各样的行业解决方案、实体化的产品等,相关产业正在逐步形成、不断丰富,相应的商业模式也在持续演进和多元化。人工智能已经成为提升国际竞争力和推进经济发展的重要科技引擎,各国正在加速人工智能产业布局与发展规划,各行业也趁着政策的东风,深耕发展人工智能,并把相应技术深入落地到现实生活的具体场景中。目前,人工智能在金融、医疗、家居、制造、零售、交通、安防、教育、航空航天等领域中都有广泛的应用。

1.4.1 人工智能在金融行业的应用

金融市场变幻莫测、风险难辨,把人工智能技术引入金融领域,不仅能推动金融业更好地服务于实体经济,提高金融服务质量,更为把控金融风险提供有力的支持。目前,人工智能技术在金融业中主要广泛应用于客户身份识别、智能客服、智能外呼、智能投顾、金融监管等场景。

1. 客户身份识别

在银行领域,当客户需要进行一些重要交易时,往往需要先进行身份的认证,避免交易错乱、盗用账户等风险事件的发生。传统的身份识别方式主要是人为核对身份证号、卡号等关键信息,效率极低。随着人工智能技术的发展,通过人脸识别、虹膜识别、指纹识别等生物识别技术快速提取客户特征进行高效身份验证的方法已推广到银行业的各大主要应用场景中,包括但不限于电子银行登录、银行柜台联网核查、ATM(自动取款机)自助开卡、远程开户、支付结算、反欺诈管理等(图1-8)。



图 1-8 客户身份识别应用



2. 智能客服

早些年，科技还不发达的时候，客户如果想咨询一些简单的银行信息、了解银行的利好产品都只能亲自跑一趟线下营业厅找专人进行解答，不仅时效性得不到保证，在柜台人员的安排和调度方面也是一种浪费。现在，以自然语言理解、知识图谱为基础的人工智能技术发展起来了，越来越多的人可以通过电话、官网、App、短信、微信、实体机器人等各种渠道与形式和银行机构进行语音或文本上的互动交流，它能快速理解客户需求，语音回复客户提出的业务咨询，并能根据客户需求导航至指定业务模块，极大地提升了银行业务办理效率，降低了银行的人员成本（图 1-9）。

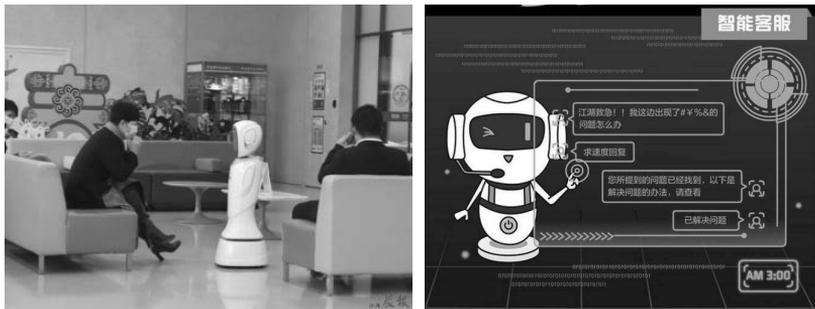


图 1-9 智能客服应用

3. 智能外呼

和智能客服类似，智能外呼基于自然语言理解、语音识别和语音合成等技术，通过电话渠道，模拟真实座席定时不定量往外呼出电话。传统的人工推销电话不仅耗费大量人力成本，收效甚微，甚至可能引起用户的反感，引得投诉率急速上升。引入智能外呼系统后，可将推销话术设定得更有趣味性，提高客户与机器人的交互欲望，从而达到相关业务的目的。目前，外呼机器人已被广泛应用于信用卡催收、通知、理财营销等各场景。

4. 智能投顾

在满足了基本的生活需求后，越来越多的人手里开始攒下多余的资金，理财的道路也由此打开。但面对海量的基金、股票、债券等理财产品，以及让人眼花缭乱的理财渠道和购买方式，人们对于精准地选择到适合自己投资偏好的理财组合的需求也日益强烈。大数据在收集到用户的社会属性、生活习惯和消费行为及投资偏好后，挖掘分析出用户画像，结合基于机器学习搭建的算法模



型给对应的人推荐出合适的投资组合，同时通过大数据及时追踪用户偏好变化、资产配置变化、产品风险变化，能动态地维护用户利益，保证利益最大化。

5. 金融监管

面对多维且海量的金融市场数据，传统的监管力度已无法满足监督要求，由遗传算法、神经网络、大数据挖掘等技术构建的智能金融数据分析专家系统，能及时监控金融市场的风险与变化，广泛应用于公司信用等级的评估、风险评估、工程管理和投资策略分析、金融和经济预测、证券价格变动的预测、破产的银行倒闭预测等领域。人工神经网络技术，是通过数据选择（数据的分离和处理）以及学习方法，对金融数据进行预测分析、管理，如股票和有价证券的预测分析、资本收益的预测和分析、风险管理以及信用等级评估等。金融数据分析遗传算法技术是借助自然选择机理的遗传算法和进化计算，实现投资交易策略的优化及管理、决策策略的优化、证券投资的选择、趋势预测模型的选择等。

1.4.2 人工智能在医疗行业的应用

随着人类平均寿命的延长，全球人口逐渐趋向于老龄化，越来越多的人需要医疗服务，但医疗资源挤兑、医护人员短缺、看病难看病贵等问题也是切实存在的，把人工智能技术引入医疗行业，将对诊疗、治疗、研发、健康管理等各方面提供全新有力的帮助。

1. 辅助诊疗

辅助诊疗主要包括电子病历和医疗影像识别。电子病历主要运用语音识别技术，可以快速帮助医生记录病人口述的病情及症状并转写成文字形式，同时利用自然语言处理技术，将非结构化病历转化为统一标准数据，方便管理和统计分析；医疗影像识别主要是通过计算机视觉及图像识别等技术，读取 CT（电子计算机断层扫描）、核磁图像等来获取患者疾病特征，同时在 AI 进行大量医疗数据的训练、学习和比对后，能快速定位患者病灶区域，辅助医生提高诊断准确率（图 1-10）。

2. 疾病预判

随着生活水平的提高，大家的消费饮食水平、生活习惯和过去相比也发生了翻天覆地的变化，但由此带来的更多不健康的饮食和生活习惯也导致了许许多多过去鲜有耳闻的重大疾病的发病率逐渐升高。医院引进人工智能技术，把机器



图 1-10 辅助医疗应用

学习、神经网络及图像识别等技术与临床数据（病灶图像、基因序列等）相结合，通过海量的疾病数据训练出相应的预判机器人模型，在下次有新的疾病数据输入时，模型能快速分析判断是否为罕见疾病，提高疾病预测的效率与准确性，为疾病的诊疗和救治争取到了更多的有效时间（图 1-11）。



图 1-11 疾病预判应用

3. 药物挖掘

药物挖掘主要应用于各类癌症的靶向药研制与发明，人工智能通过深度学习构建的神经网络，提取挖掘文献资料及疾病图片中的相关数据信息，利用递归神经网络、LSTM（长短期记忆网络）和卷积神经网络对相关信息进行处理整合，分析各类有效靶向药及病理之间的相关性，挖掘出新的靶点，再通过大数据分析找到符合此类靶向药物的病人，优先进行临床试验，为新型药物的验证提供了很大帮助（图 1-12）。



图 1-12 药物挖掘应用

4. 智能健康管理

通过智能可穿戴设备对个人日常活动和生理体征参数等健康数据采集，通过大数据分析能有效评估病人整体状态，及时预判疾病风险，规划日常生活饮食，定制健康管理计划，同时还能结合定期体检信息，完善健康评估和健康管理服务的建立；此外，大部分的智能健康管理软件还会嵌入虚拟的 AI 医生，对常见的、简单的医疗问题进行解答，提供远程医疗服务，从而实现个人健康风险的有效管控（图 1-13）。



图 1-13 智能健康管理应用

1.4.3 人工智能在家居行业的应用

随着社会的高速发展，人们在职场的压力也越来越大，回到家里后都想马上放松身心，享受最惬意的高质量生活。在家电控制、家居生活等场景中引入人工智能技术，把各类家电联结起来，形成一个庞大的物联网（internet of things, IoT），可以给人们带来极大的舒适与便利。



1. 家电控制

家电控制指通过语音识别、手势控制、指纹、声纹及面部识别技术对家电进行开启、关闭、调整等动作。例如，智能音箱的唤醒、点歌、闹钟功能、音量调整，主要是通过语音识别及语义理解技术，把人类说出的话转化成 AI 可识别的计算机语言，让 AI 获取到人类的意图，再执行相关的操作；智能门锁则可以通过提前录入的人类的声纹、指纹、面部特征等，运用计算机视觉等技术，识别该房子正确的主人，当摄像头记录到陌生人员在门前逗留过久时，还可以及时记录并向主人发送警报信息，提高安全性（图 1-14）。



图 1-14 家电控制应用

2. 家庭机器人

目前日常家居中最常见的家庭实体机器人主要为扫地机器人和拖地机器人。它们利用计算机视觉和红外线扫描技术，能自动地扫描家庭路线及路线上的障碍信息，自动规划避障路线，通过定时的打扫任务，自动进行智能清扫；在遇到楼梯、家居和动物时，也能及时停止前进，防碰撞防跌倒；此外，市面上大部分的智能机器人都可以与手机 App 相连，通过手机可查看它们的清扫轨迹，并及时操控它们进行自清洁、自烘干、自充电等操作，非常便捷（图 1-15）。

3. 智能厨房

目前在厨房领域，人工智能的应用大部分还停留在较为基础的阶段，如手势操控抽油烟机开启、定时的智能电饭煲等。但也有部分厂商正在尝试研制搭载语音交互技术的智能冰箱，这种冰箱能自动识别冰箱内的物体，区分快过期/未过期的蔬果，及时提醒人们进行清理；收集人类的饮食习惯、健康数据，进行个性化菜谱推荐等功能也在此行业的研究范围中，期待技术上能尽早攻克此类研究的难题，为人类的生活提供更大的便利（图 1-16）。

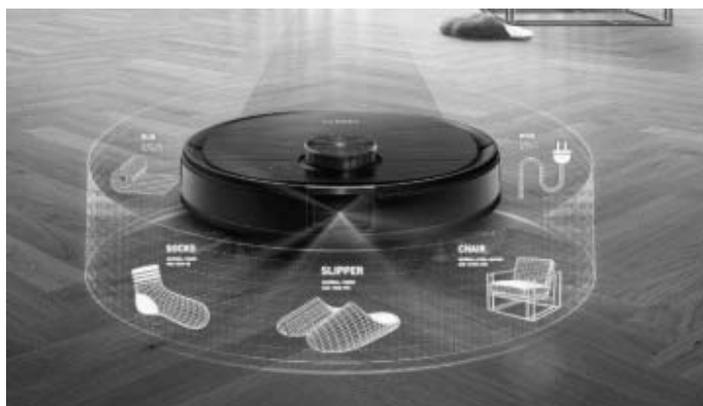


图 1-15 家庭机器人应用



图 1-16 智能厨房应用

1.4.4 人工智能在制造行业的应用

科技时代背景下，传统制造行业单纯依靠人力和重复性机械劳动的运作模式、落后低效的制造工艺等已经不再适用于当下，面对招工难、成本高、利润低、效率差的困境，越来越多的企业开始寻求转型创新，在建设数字化工厂的道路上探索前进。引入人工智能技术，将在制造、检测、维护等各方面为制造行业带来颠覆性的改革。

1. 智能产品检测

采用基于深度学习的计算机视觉技术，把 AI 与缺陷检测结合，通过多次对产品缺陷图片的输入与学习，建立全套的缺陷检测系统，能快速地替代人工



检出不合规范的产品，一方面统一了检测的标准，避免人为主观带来的影响；另一方面也为企业节省了大量的人工成本。目前该技术已经广泛应用于面板检测、电路板检测、纺织品检测等工业领域。

2. 智能制造

在实际的生产制造过程中，工艺程序可能会受到温度、湿度、环境整洁度等多重因子的影响，而人为模拟很难制造出百分之百的理想环境，这时我们引入智能生产的思路，利用机器学习建立模型，把内外部的参数影响转换成计算机语言，多次的动态调节和参数修正，让产品的设计和优化能自动化地完成全过程，从而也保证整个过程中最大限度地避免其他因素的干扰，对于生产设计有很大的帮助。

3. 智能设备维护

工业生产中往往会使用到大量的重型机器人、机械手臂，而对于这些冷冰冰的机器，一颗小螺丝的掉落都有可能引发大批量的生产宕机或良率损失，严重的情况还会造成人身安全事故的发生，故周期性的点检和维护就显得尤为重要。这时如果我们引入智能维护系统，通过每日采集机组各部位的振动数据、润滑数据、温度变化数据，输入我们构建好的机器学习模型，将其输出的数值与真实运行数值进行对比，及时作出自动告警和提醒，能有效监测设备寿命，并提高整个系统的安全性。

1.4.5 人工智能在零售行业的应用

从 2005 年开始，网络购物风潮不断兴起。2019 年底，新冠肺炎疫情开始肆虐发展。受两者影响，世界各地的实体经济都受到不小的打击，越来越多的零售企业在开展线下运营时，不得不重新评估经营方式。通过人工智能技术加以辅助判断，会为商家带来新的管理模式，从而达到降本增效的目的。

1. 智能仓储

智能仓储主要包括智能仓库选址、智能库存管理及智能分拣等。智能仓库选址是指收集客流、供应商位置、生产商位置、运输成本、劳动力、建筑成本、税收制度等数据，利用大数据分析做出最优的选址方案，降低企业成本和人为主观的干预；智能库存管理是指结合历史的顾客消费数据、分布地区等，采用深度学习、宽度学习等算法，构建相关的需求量预测模型，通过不断的数据输



人和动态变化,形成一个智能的仓储预测系统,相应地动态调整各区库存数量,有助于企业高效地进行库存调配,避免资源浪费;智能分拣也是 AI 在仓储管理中的一个重要应用,它的体现形式主要为仓库中的穿梭及分拣机器人,系统通过判断货物的重量、外观、属性、目的地等,先将它们大致做一个分区,而分拣机器人通过图像识别货物的快递面单信息,根据货物的目的地进行分拣和运送,大大地提高了分拣速度和正确性(图 1-17)。



图 1-17 智能仓储应用

2. 无人商店

无人商店利用人脸识别、视频识别等技术,结合压力传感器及红外探测的应用,可以实现无人经营、自动结账的运营模式。用户先注册绑定商店的账号并录入相应的人脸特征信息,当步入商店的时候,摄像头自动捕捉识别人脸并关联该用户账户,再通过货架上的摄像头、传感器等装置,判断用户在什么位置拿走了什么商品,自动进行商品的结算和消费,避免了长时间排队的等候,提高了消费体验和便捷性,也为企业节省了人力成本(图 1-18)。



图 1-18 无人商店应用



1.4.6 人工智能在交通行业的应用

当前的时代背景下，国家大力推进智慧城市^①建设，智慧交通是其中一个重要组成部分。现代的城市交通系统不再是比较谁的马路更宽，而是看交通智能化的程度有多高。人工智能技术和交通运输系统融合后，不仅能为基础交通设施建设、运输装备研发、运输服务等方面提供有力帮助，还能有效地提升整个城市的交通枢纽运行效率，保障居民出行安全，降低运输成本。

1. 交通路口信号灯优化

红绿灯信号系统采用模糊控制、遗传算法、神经网络等基础的人工智能核心技术，结合蚁群算法、粒子群优化算法等，通过计算机视觉技术分析摄像头拍摄到的人车排队拥堵情况，结合路上人车流量及其他路口交通灯情况，动态调整交通灯的切换时间，把路口的控制模式由车等灯转变到灯看车，扩大了监控面与监控场景，提高了红绿灯切换效率，从而优化交通信号系统（图 1-19）。



图 1-19 交通路口信号灯优化应用

2. 高速路收费稽查/电子不停车收费

高速路有很多路口，不同的路径、不同的车型，收费是不一样的，只靠人工的监管和排查无法完全规避违法人员通过换车牌、换车头来逃费的行为。ETC（电子不停车收费）系统采用 Python+OpenCV 人工智能组合识别车辆信息并预判车辆即将驶入的车道，通过“车脸”识别，对车进行全程路径跟踪，及时扣费，避免偷逃费；同时利用激光雷达探测及短程通信技术，不断扫描激光雷达与被测车辆间的距离，计算出车辆到达 RSU 设备^②有效识别范围的时间；当车

① 智慧城市是指利用各种信息技术或创新概念，将城市的系统和服务打通、集成，以提升资源运用的效率，优化城市管理和服务，以及改善市民生活质量。

② RSU（路侧单元）设备是 ETC 系统中安装在路侧，采用短程通信技术与车载单元进行通信，实现车辆身份识别，电子扣分的装置。



辆进入 RSU 中天线设备的识别范围时, RSU 中天线设备发出无线电信号, 进而实现 OBU^①设备与 RSU 设备的通信。利用以上人工智能技术加上传统的计费系统, 就可以实现不停车收费(图 1-20)。



图 1-20 高速路不停车收费应用

3. 车路协同

随着科技的发展, 越来越多的事物依赖于网络技术而生。据国家发布的车联网和车路协同的规划, 未来要求所有车必须联网: 车辆上将安装通信模组、CPU (中央处理器) 芯片、AI 芯片、传感器、北斗导航等设备, 路面上也会安装各种传感并连接网络。通过高精定位及高精地图、场景算法、车辆自组网、DSRC^②、C-V2X^③等, 越来越多的车辆和道路会联网进行信息沟通与交换, 道路的环境情况会实时通知给车辆, 车结合自身的环境识别能力作出行驶判断, 从而实现车路协同(图 1-21)。

4. 无人驾驶

很多人第一次听到无人驾驶技术可能是在动画片《哆啦 A 梦》中, 但其实在 20 世纪 80 年代初, 美国国防高级研究计划署就已经开始与陆军合作发起自主地面车辆 (ALV) 计划, 并多次举办无人驾驶挑战赛; 2009 年, 谷歌公司宣布组建人工智能团队开始研发无人驾驶技术; 从 2013 年开始, 奥迪、沃尔沃、宝马、特斯拉等国外的知名传统汽车厂商纷纷开始布局无人驾驶汽车产业。随

① OBU (车载单元) 是安装在车辆上可以与 RSU 进行通信的装置。

② DSRC 即专用短程通信技术, 是一种专门用于机动车辆在高速公路等收费点实现不停车自动收费 ETC 技术。

③ C-V2X (蜂窝互联网) 是一种基于蜂窝网络的车用无线通信技术。

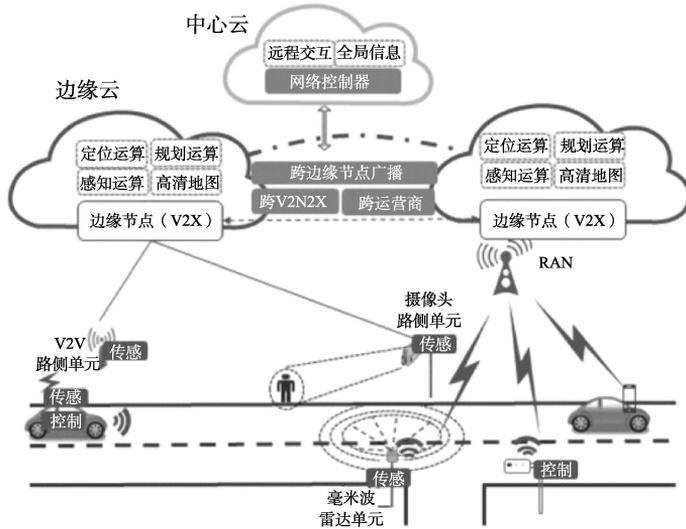


图 1-21 车路协同应用

着丰田 2021 年在日本发售全球首款获法律许可的 L3^①自动驾驶量产车,越来越多的车企开始实现高级别自动驾驶规模化量产。国内几乎也是在同时间段进行无人驾驶技术的研究,1992 年,国防科技大学成功研制出中国第一辆真正意义上的无人驾驶汽车;2015 年,百度宣布正式成立自动驾驶事业部,同时,一汽、上汽、北汽、奇瑞、长安等国产汽车厂商也加快与国内高校合作研发无人驾驶技术,争取尽快实现自动驾驶汽车的商用化及量产;2020 年,中国长沙等多个城市已经开始了 L3 的无人驾驶出租的试运营,深圳也开始试验运行无人公交。

以上几种无人驾驶的试验运行模式主要是通过摄像头、激光雷达、毫米波雷达、超声波传感器来感知环境,结合行车电脑判断环境路况(涉及图像语义分割、目标检测、立体视觉匹配),均为有安全员的、按照指定路径的无人驾驶。想要实现完全的无路线无规划的 L5^②无人驾驶,还需要人工智能研究者更多的努力和实践(图 1-22)。

5. 违章抓拍

违章抓拍是人工智能在交通行业应用得较为广泛和成熟的场景,它的工作

① L3 级别的无人驾驶模式是有条件的自动化,指自动化系统完成大部分的驾驶操作,但当紧急情况发生时,驾驶员需作出相应的干预。

② L5 级别的无人驾驶模式是全自动驾驶,全程无须人类干预。



图 1-22 无人驾驶应用

原理是利用车牌识别、车违章行为识别和抓拍技术，对违章进行抓拍罚款。通过道路的电子眼进行机动车图片抓拍，采用计算机视觉技术进行车辆号牌识别、车辆速度检测、布控比对报警、查报站出警拦截，其中涉及车辆图像采集（视频识别、地感线圈识别）、车牌定位（边缘检测定位、颜色定位、特征工程定位、神经网络定位）、车牌字符分割（基于连通域标记的算法、基于字符几何特征的算法、基于图像投影的算法）、车牌字符识别（模糊匹配、神经网络、支持向量机）等技术。

6. 自动停车场

目前人工智能技术在自动停车场的主要应用是通过高清摄像头进行车牌识别，加上计时系统，在闸机卡扣实现自动计费收费。自动收费系统很大程度上缓解了人工收费的压力，并且有效地降低了停车场出入口的拥堵频率。它主要涉及的技术包括车辆图像采集（视频识别、地感线圈识别）、车牌定位（边缘检测定位、颜色定位、特征工程定位、神经网络定位）、车牌字符分割（基于连通域标记的算法、基于字符几何特征的算法、基于图像投影的算法）、车牌字符识别（模糊匹配、神经网络、支持向量机）、计费设置、移动支付等。在未来智慧城市的规划中，停车场将对接城市大脑^①，可将停车场实时车位进行同步，并给司机推荐最近的空闲停车场（图 1-23）。

^① 城市大脑可以理解为城市的大脑，它将散落在城市各个角落的数据汇聚起来，搭建一个用云计算、大数据、人工智能等前沿技术构建的平台型人工智能中枢，是一个对城市信息进行处理和调度的超级人工智能系统。



图 1-23 自动停车场应用

1.4.7 人工智能在安防行业的应用

随着智慧城市^①建设进程的加快，传统的安防手段已经不能满足各行业的监控需求。面对海量的监控图片、监控视频及密集的流动人口，采用人工智能技术，能协助工作人员快速开展安防工作，提升监控效率。

1. 智慧警务

在公安行业中，违法犯罪的案件不仅多，而且杂。面对海量的犯罪嫌疑人信息，使用人为方法一个个地进行查看辨别无疑是巨大的工作量，公安民警迫切需要在海量的视频信息中快速发现犯罪嫌疑人的线索。依托于信息感知、云计算、人工智能等技术的不断发展，公安部门正在大力推进公安信息化及智慧警务建设，人工智能在其中发挥着越来越重要的作用。尤其在视频内容的特征提取、内容理解方面，人工智能有着天然的优势。基于深度学习的图像识别技术在对人、车、物进行检测和识别的过程中发挥着重要作用。公安机构的身份管理系统运用人脸识别技术在布控排查、犯罪嫌疑人识别、人像鉴定以及重点场所门禁等领域获得了良好的应用效果；同时，治安监控系统也搭建了对应的人员管理数据库，将案件重点关注人群入库，对系统进行智能化升级，实现对常住人口、暂住人口、重点人口、在逃人员等人群的人像比对，为户籍管理、治安管理、刑侦破案等提供大数据分析技术手段（图 1-24）。

^① 智慧城市是运用信息通信技术，有效整合各类城市管理系统，实现城市各系统间信息资源共享和业务协同，推动城市管理和公共服务智慧化，提升城市运行管理和公共服务水平，提高城市居民幸福感和满意度，实现可持续发展的一种创新型城市。



图 1-24 智慧警务应用

2. 智慧社区

社区是城市的基本空间，是社会互动的重要场所。伴随着人口流动性加大，社区中人、车、物多种信息重叠，数据海量复杂，传统的人工管理方式难以实现高效的社区安防管控，同时，社区管理与民生服务息息相关，不仅需要在管理上实现技术升级，还要实现大数据下的社区服务。通过社区监控系统中融入人脸识别、车辆分析、视频结构化算法等技术，对有效视频内容进行提取，不但可以检测运动目标，还可以根据人员属性、车辆属性、人体属性等多种目标信息进行分类。与公安系统结合后，还可以分析犯罪嫌疑人线索，为公安办案提供有效的帮助。另外，在智慧社区中使用带有人脸识别功能的智能门禁等产品也能够精准地进行人员甄别（图 1-25）。



图 1-25 智慧社区应用



3. 疫情防控

新冠肺炎疫情发生以来，防控形势严峻，疫情的排查仅凭人力收效甚微，这时，人工智能技术在信息收集、支持复工复产等诸多方面就发挥出显著的作用了。在具体实践中，智能服务机器人、大数据分析系统和智能识别（测温）产品数量居前三，计算机视觉和智能语音等人工智能技术成熟度相对较高，面向的场景丰富，在抗击疫情中也发挥了极大作用。

首先，通过使用智能外呼场景，对途经重点地区、涉疫风险地区的人员进行智能外呼，收集相关信息，大大地提高了筛查效率，减轻了基层工作者的走访摸排压力；其次，实体智能服务机器人可以完成智能清洁、消毒和配送等重复机械的简单工作，降低了人员感染风险；再次，大数据分析系统则可以为医院、疾控中心、政府机关、企业、社区及群众提供疫情地图、人群追踪、同乘查询、趋势预测与舆情分析等服务，便于灵活调整防疫政策。另外，广泛应用在地铁、车站、机场等关键交通枢纽及大型企业、商场等入口的智能识别（测温）产品，可以实现多人同时非接触测温、体温异常报警、人脸识别，并对数据进行实时上云、跟踪管理（图 1-26）。



图 1-26 疫情防控应用

1.4.8 人工智能在教育行业的应用

随着群体素质的提高和义务教育的普及，越来越多的人开始意识到教育的重要性。但各地教学资源的不均衡是教育行业的难点之一，在这种情况下，互联网教育模式应运而生。教育教学过程中引入人工智能技术，形成模式化的学习和分享过程，给教育行业开辟了一条新的道路，为实现信息化教育提供了极大的便利，也更充分合理地利用了有限的教育资源。

1. 口语评测和口语对话

顺应国际化发展的趋势，英语在国内得到了快速的普及与应用，越来越多



的学习者对英语的需求已经不仅仅停留在会听会看，而是更注重听说读写的全面发展，而在“说”这个方面，英语专家的口语指导必不可少。人工智能在口语教学方面的应用包括口语评测和口语对话。口语评测包含三类：朗读与复述、陈述与讨论、演讲与问答，其原理是通过 ASR 技术对学生的发音进行转写，然后通过 NLU 进行打分。目前很多英语学习的 App 有跟读和配音的功能，且设置了丰富的对话场景来考查学习者的口语能力；此外，部分 App 还开发了普通话等级测评功能，它拥有固定的测试对比材料，主要针对学习者的音调、语速、准确清晰度等进行识别和考核。口语对话则是在此基础上根据预先设定的脚本，在语义理解后进行匹配对应的对话脚本，把应该回答的对话内容通过机器人，采用 TTS 技术进行播报，从而达成一问一答的对话效果，使口语训练得到有效提升。

2. 智能搜题和智能批改

智能搜题和智能批改相信家长与老师都不陌生，都是 OCR（光学字符识别）的文字识别和手写识别的应用。目前 OCR 技术已趋于成熟，手写体识别准确率可达 90%以上，印刷体的识别准确率更高。智能搜题和智能批改的核心原理是类似的，都是先通过 OCR 技术将拍照后得到的题目和学生的答案进行转写，再通过 NLP 对识别的题目进行答案匹配或者对识别的答案进行客观打分。智能搜题不再单纯依靠纸质的答案或者老师的讲解，能有效地提升学生的课外辅导效率。智能批改也可以大大地解放老师的双手和工作量，提升教学的效率(图 1-27)。



图 1-27 智能搜题和智能批改应用

3. 智慧课堂（虚拟课堂）

为了在课后给学生提供更多的辅导，市面上的线上教育 App 也不断推陈出新，开发出了虚拟课堂的教育方式。它主要包括对话方式实时反馈、个性化对话辅导、课堂专注度分析等功能。前两个功能采用了语音交互和测评技术，对于学生的回答、朗读进行识别、评分，并可以通过 TTS 技术，和学生形成交互。



对话方式实时反馈通过建立虚拟教室实时跟学生对话，让学生感觉更亲切、接受度更高。个性化对话辅导则是通过分析学生的学习目标、学习数据、学习反馈等，利用算法计算出相关模型，针对不同的学习者提供个性化的辅导方案。课堂专注度分析通过计算机视觉抓拍学生上课时的照片、视频，捕捉到学生的神情、表情、面部姿态、行为、面部角度等，实现抬头率、看手机率、微笑率、专注度、学生人数、离席率等课堂效果指标的智能统计，从而对学生课堂状态进行自动监测（图 1-28）。

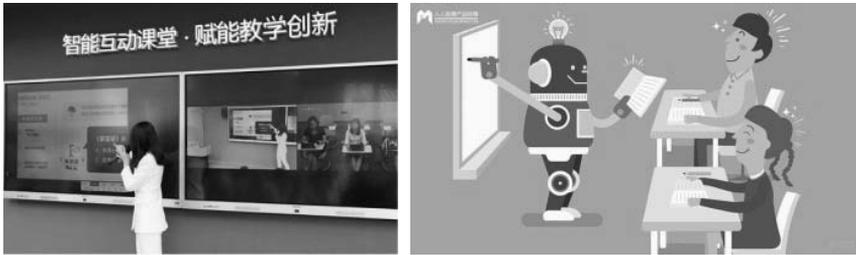


图 1-28 智慧课堂应用

4. 个性化教学

随着 AI 教学的推进，学习者与机器人的互动变得频繁，可收集和利用的数据越来越多。根据人工智能算法和大数据分析，在积累一定量的教学数据后，可以由文本分析引擎推算出更适合某一学生的课程学习计划和练习题。例如将同一门学科学习得较差的学生归为一类，并判断该类学生对学科的掌握程度、认知水平、个性特征及学习方式，为其制订相应的课程学习计划，从而使该类学生接受知识的效率大大提升。

1.4.9 人工智能在航空航天领域的应用

航空航天领域对于安全性和精密性都有着极高的要求，单纯靠人力是无法完成精准的计划和安排的。在这里，人工智能技术的科技性和严谨性可以得到有力的发挥。

1. 航空航天维修智能化

都说飞机是世界上失事概率最低的交通工具，但只要发生事故，机上人员的存活概率基本为零。由此可见对飞机的定时维修、故障预测极为重要，越来越多的航空公司也开始寻求 AI 技术来对飞机进行预测性的维修。通过对飞机上广泛分布的传感器上传的数据进行分析，可以识别和报告潜在故障，并预测



最合适的维修时间，从而创建更智能的维修计划。

2. 智能排班

航空公司运行及飞机执行过程中，飞机的飞行排班及相关机务人员的排班无疑也是重要环节之一。一个航空公司每天需要执飞世界各地、各个时间段的航班，如何调配飞机航线、起飞时间、机舱人员，单靠人为的计算和安排是远远不能实现的。人工智能神经网络、遗传算法可基于丰富的航班运行动态、机场状态等海量信息，通过精密的计算，优化机组人员排班，以解决手工排班效率低、排班的有效工时低、各处室分工不均等问题，为航空公司带来极大的管理效率提升，并避免潜在的经济损失。